Introduction to SNPs and STRs

Single Nucleotide Polymorphisms and Short Tandem Repeats

UZH-BIO392
Feifei Xia
feifei.xia@uzh.ch

Introduction to SNPs

Single Nucleotide Polymorphisms

Types of Human Genetic Variants

Differences in DNA among individuals drive many types of phenotypic differences.
There are many types of genomic differences between individuals

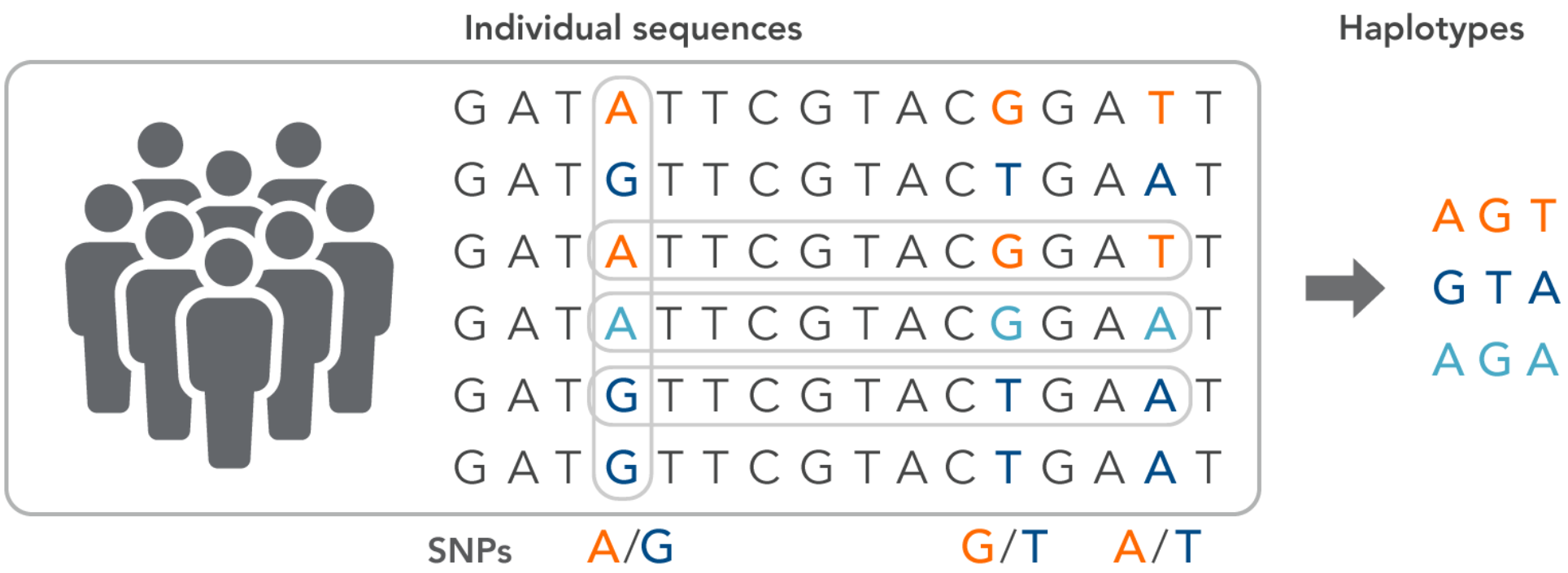
- Structural variations (SVs)
- Copy number variations (CNVs)
- **Short tandem repeats (STRs)**
- **Single nucleotide polymorphisms (SNPs)**

What are SNPs?

- SNPs are the most common type of genetic variation in the human genome. They represent a single base change (A, T, C, or G) at a specific position in the DNA sequence
- We distinguish between SNPs which are relatively common (occurring in at least 1% of the population) which are called single nucleotide polymorphisms or SNPs, and rarer variants called SNVs.

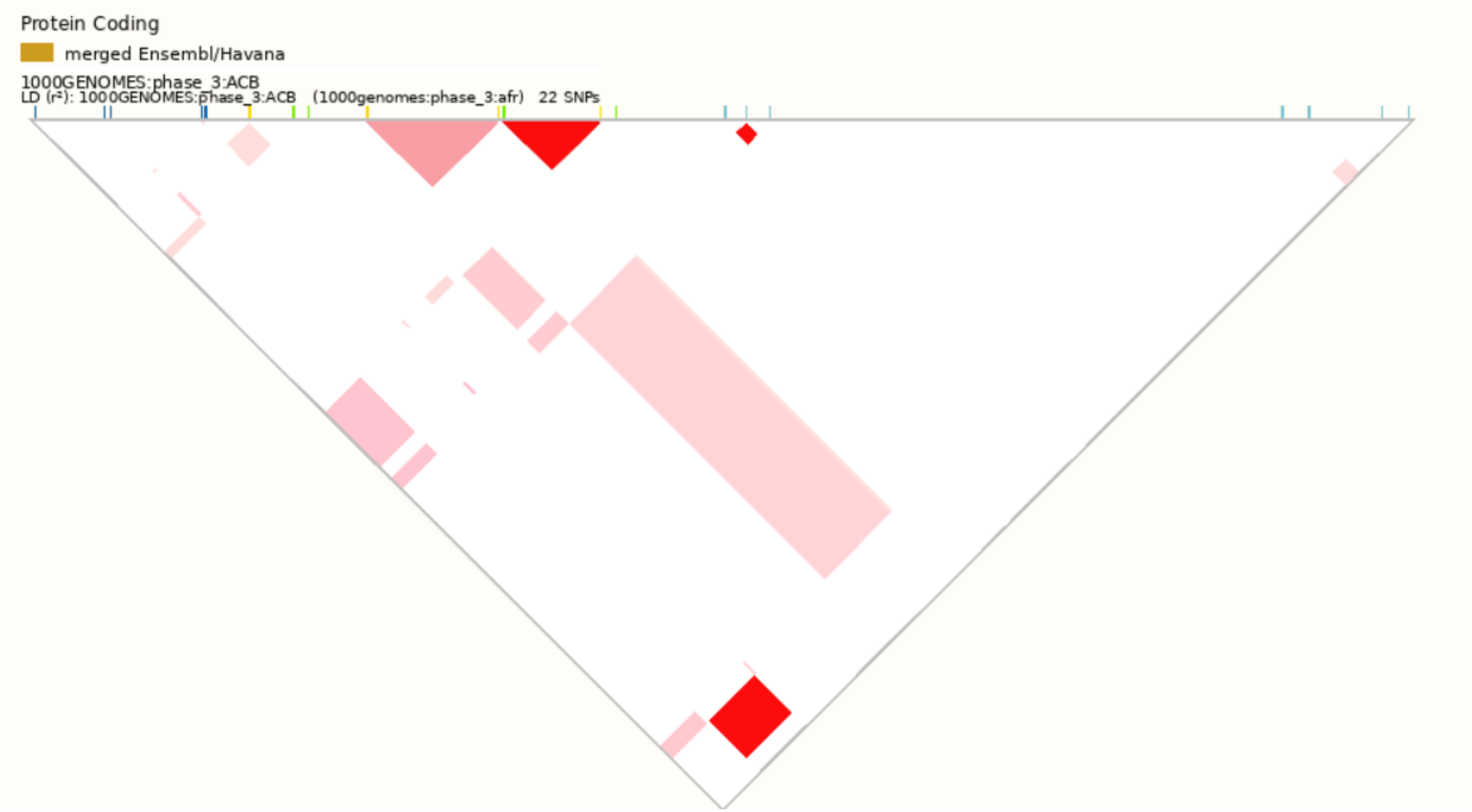
SNP Haplotype

- A SNP haplotype is a set of closely linked SNPs on the same chromosome that are inherited together as a block from a single parent.
- It represents a specific combination of alleles at multiple SNP positions.



Linkage Disequilibrium

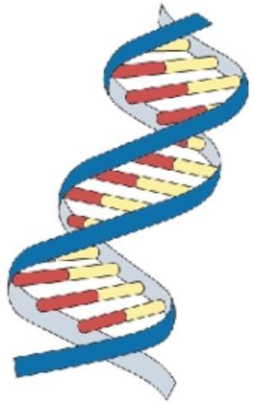
LD refers to the non-random association of alleles at different loci in a population.

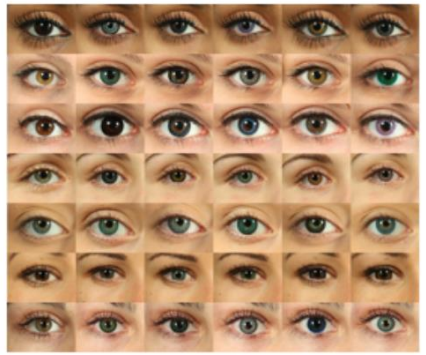


https://www.ensembl.org/Help/View?id=197

Genotypes to phenotypes

- Most SNPs have no effect on health or development. However, some of them have proven to be very important in the study of human health.
- Genetic variation between individuals leads to differences in an individual's **phenotype**, trait or risk of developing a disease. An individual's phenotype is influenced both by their genotype and their environment



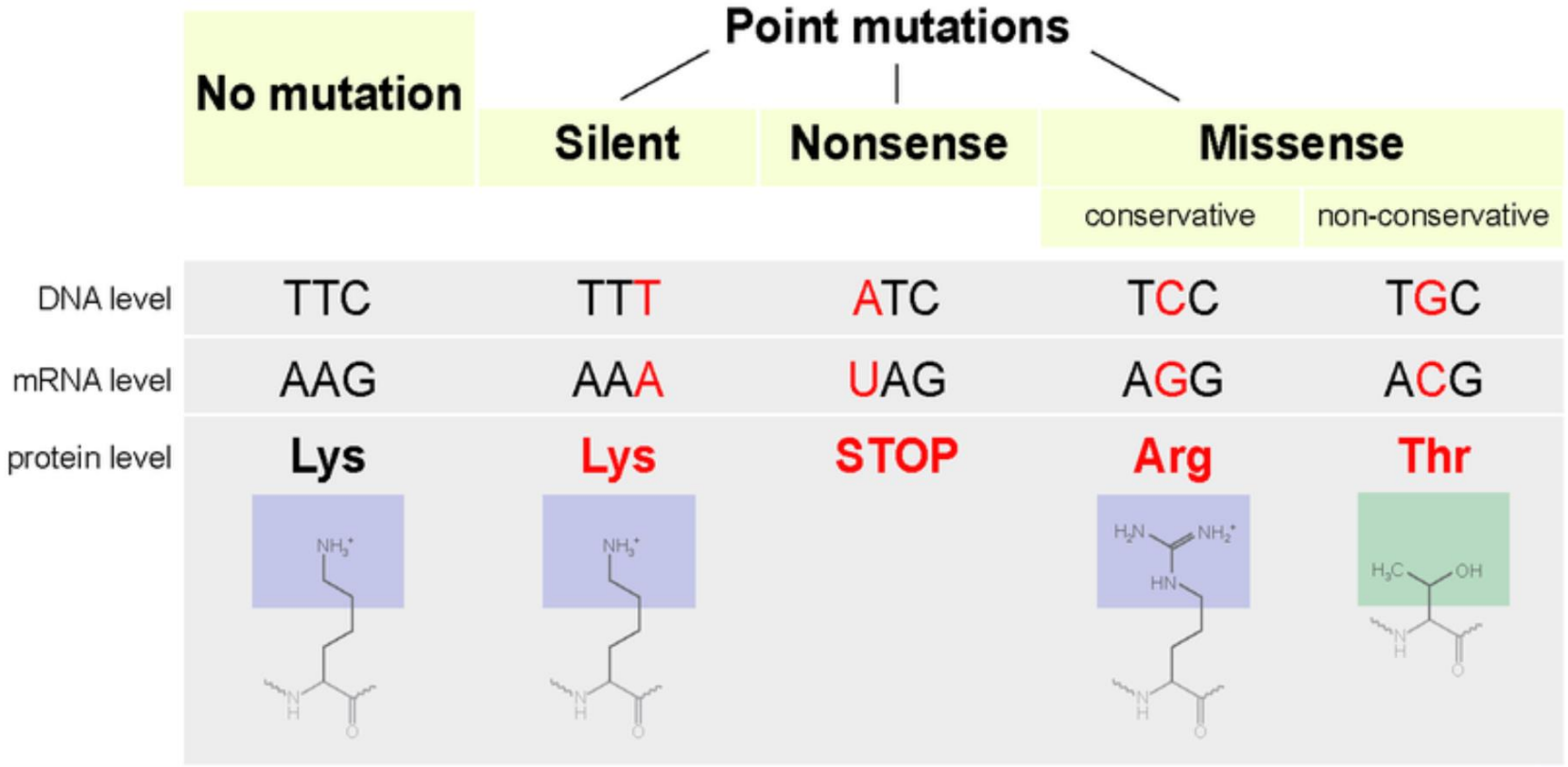


Genotypes are the genetic make-up of an individual.

Phenotypes are the physical traits and characteristics of an individual and are influenced by their genotype and the environment.

What effect do SNPs in coding regions have?

When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene's function.



SNP Genotyping Methods

SNP genotyping are techniques that analyze genomic sequence variations. These single-base substitutions are typically detected using real-time PCR, microarrays, or next-generation sequencing (NGS) techniques.

Real-Time PCR

- High sensitivity
- Familiar workflow
- Capital equipment available in most labs

Microarrays

- Trusted data quality and proven track record
- Scalable and cost-effective for detecting known SNPs in large population studies
- Can focus on target-rich areas
- Robust data analysis pipeline and uncomplicated variant interpretation

NGS

- Unbiased variant discovery using only limited knowledge of the presence or nature of the variants
- High-throughput capabilities and scalability
- Small sample size input
- Emerging user-friendly tools to simplify data analysis

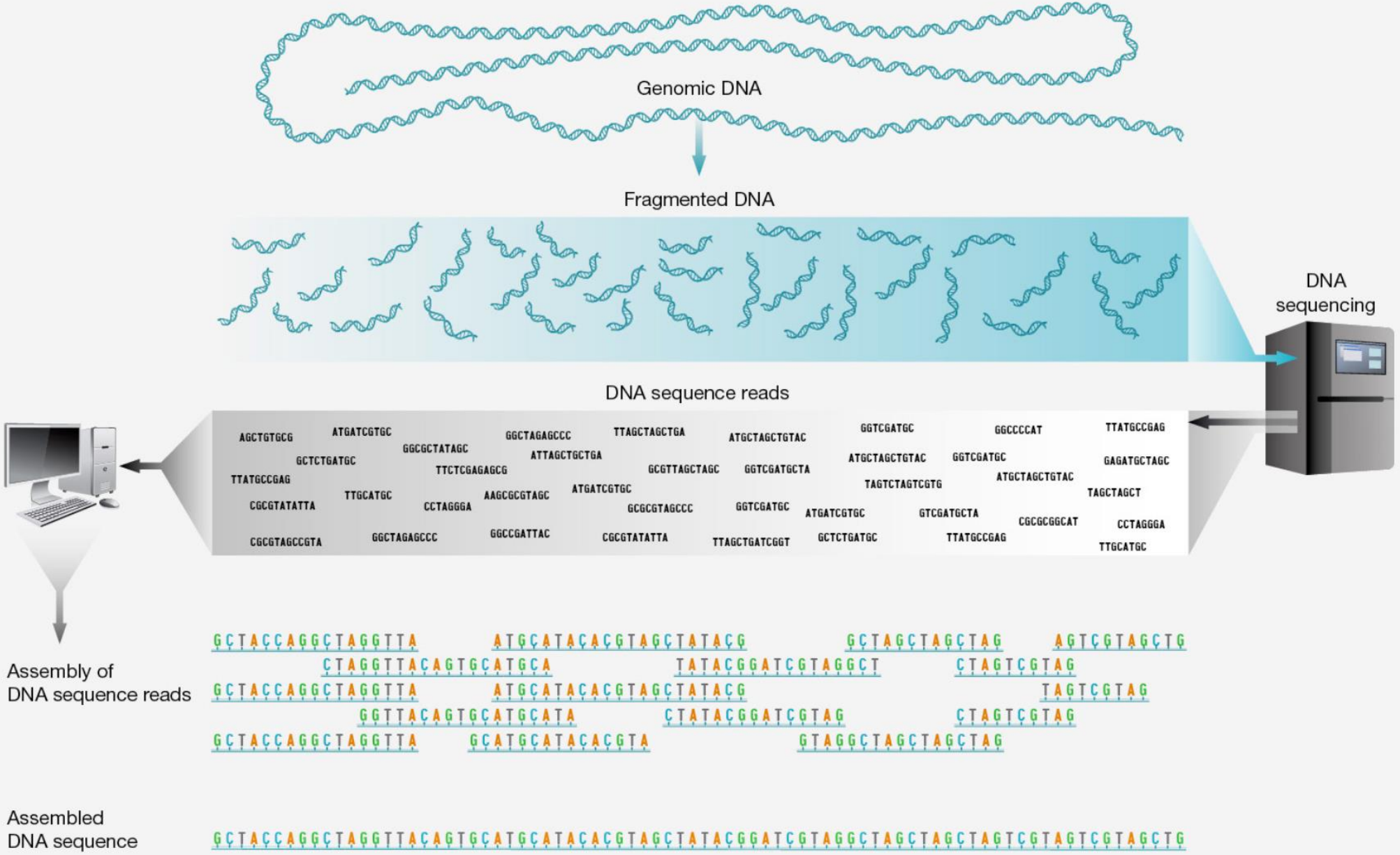
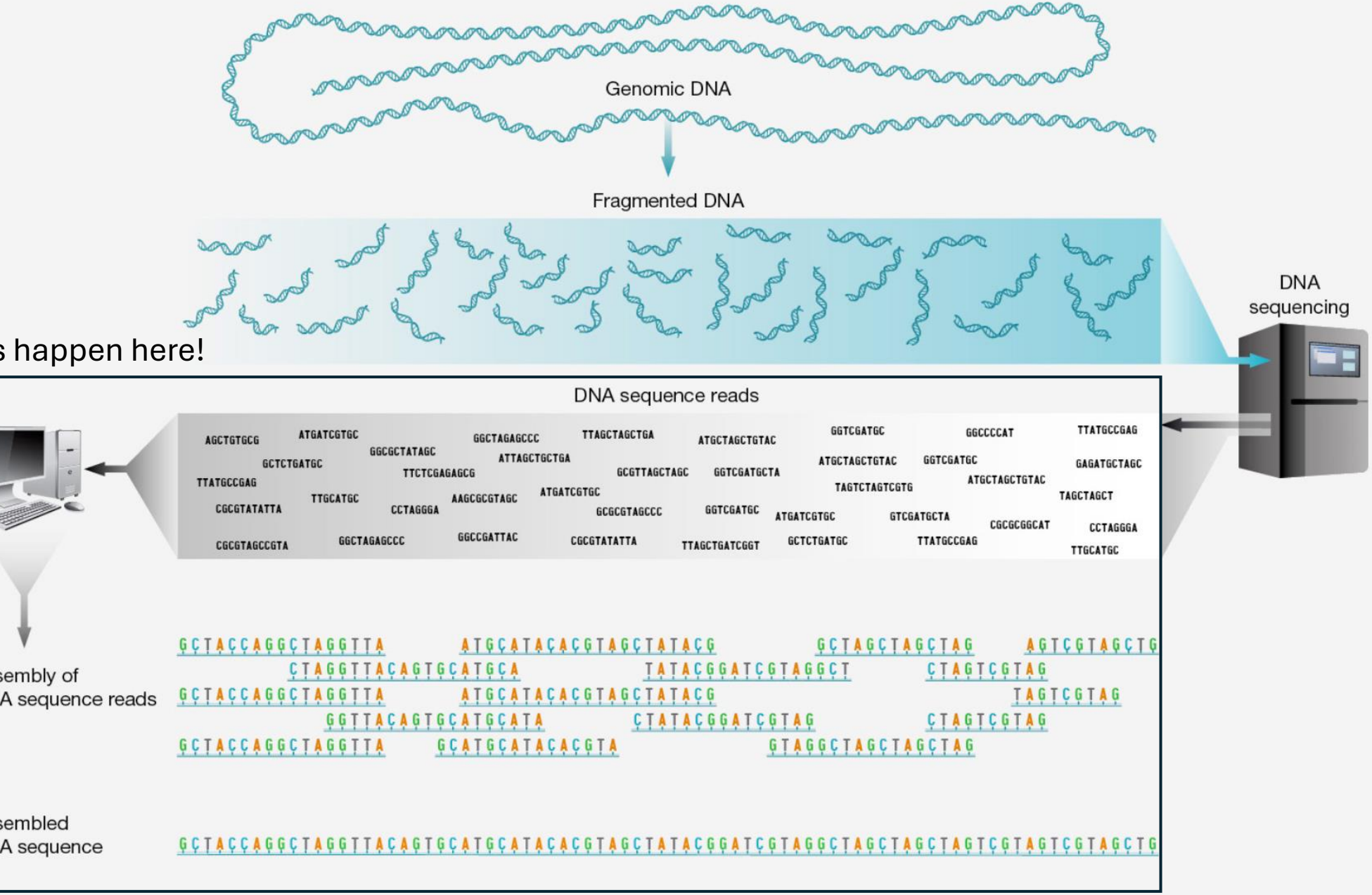
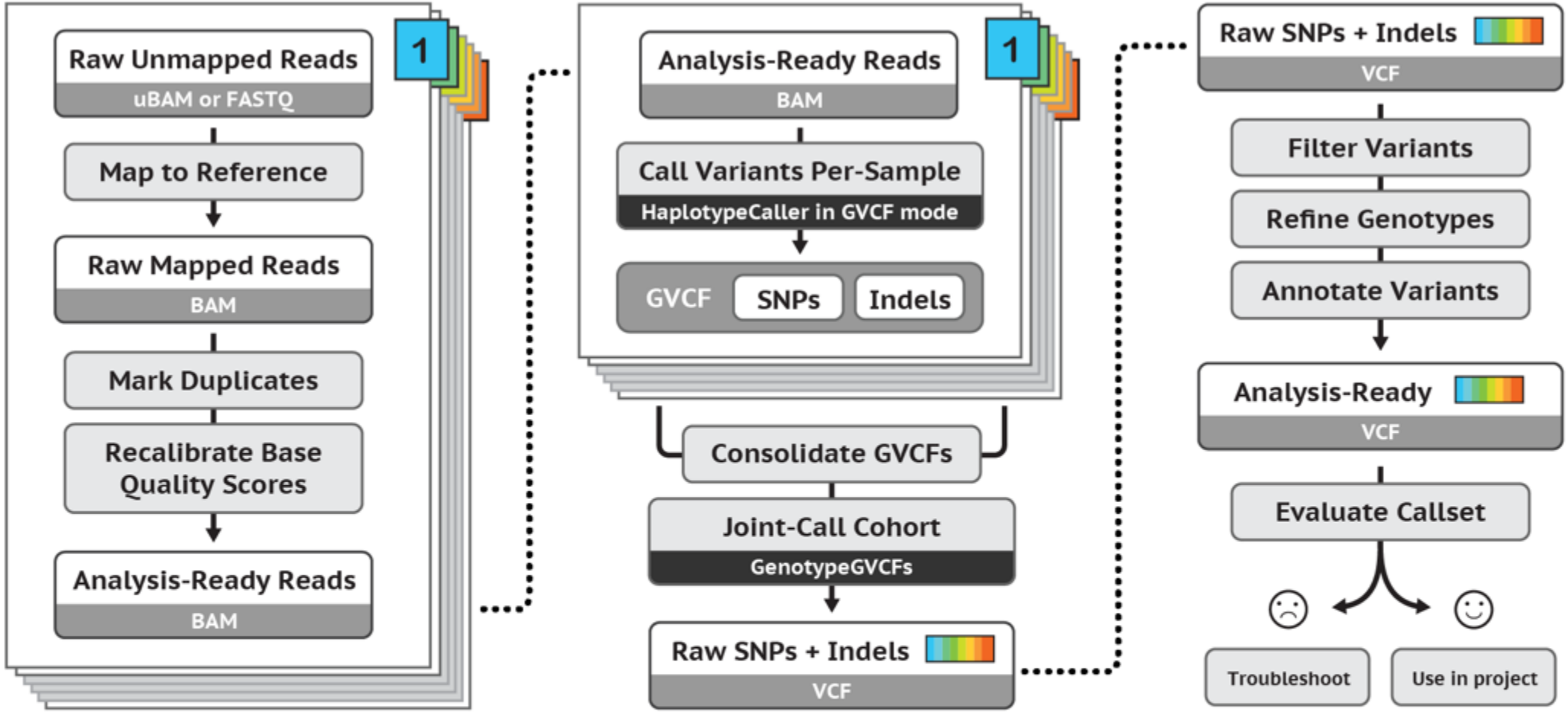


Image courtesy of National Human Genome Research Institute (https://www.genome.gov/genetics-glossary/Shotgun-Sequencing)



Bioinformatics happen here!

Image courtesy of National Human Genome Research Institute (https://www.genome.gov/genetics-glossary/Shotgun-Sequencing)



Best Practices for SNP and Indel discovery in germline DNA - leveraging groundbreaking methods for combined power and scalability.

https://gatk.broadinstitute.org/hc/en-us

Filtering SNPs

To reduce the number of markers, filtering is often done to remove uninformative SNPs.

- remove SNPs with poor quality
- remove SNPs with identical calls in all the samples
- based on biological functioning.

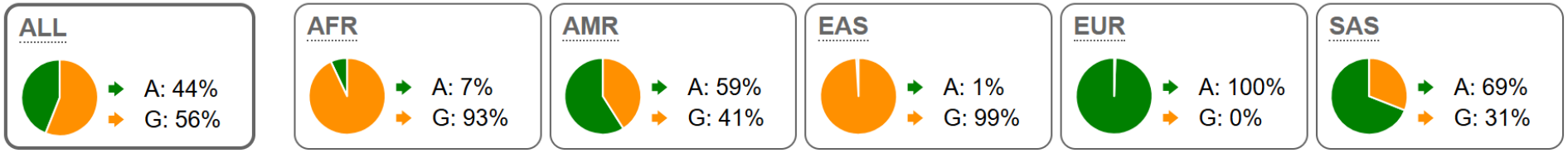
Applications of SNPs

- Genetic Ancestry and Population Genetics
- Genome-Wide Association Studies (GWAS)

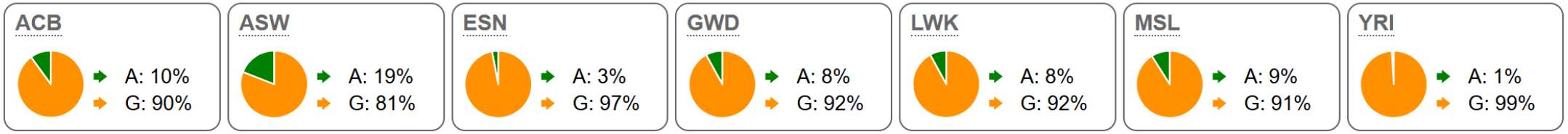
Population Structure

Population genetics

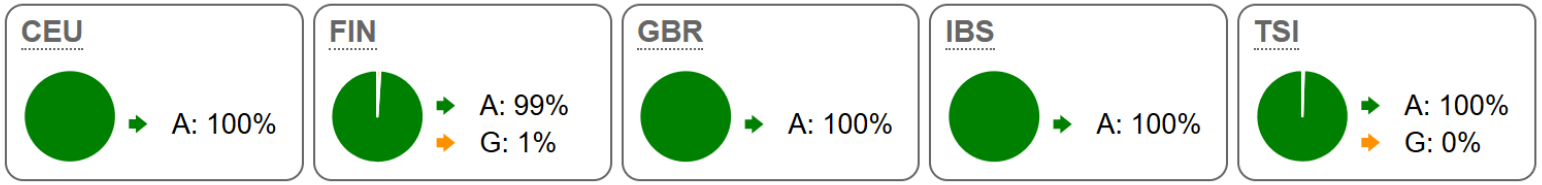
1000 Genomes Project Phase 3 allele frequencies



AFR sub-populations



EUR sub-populations

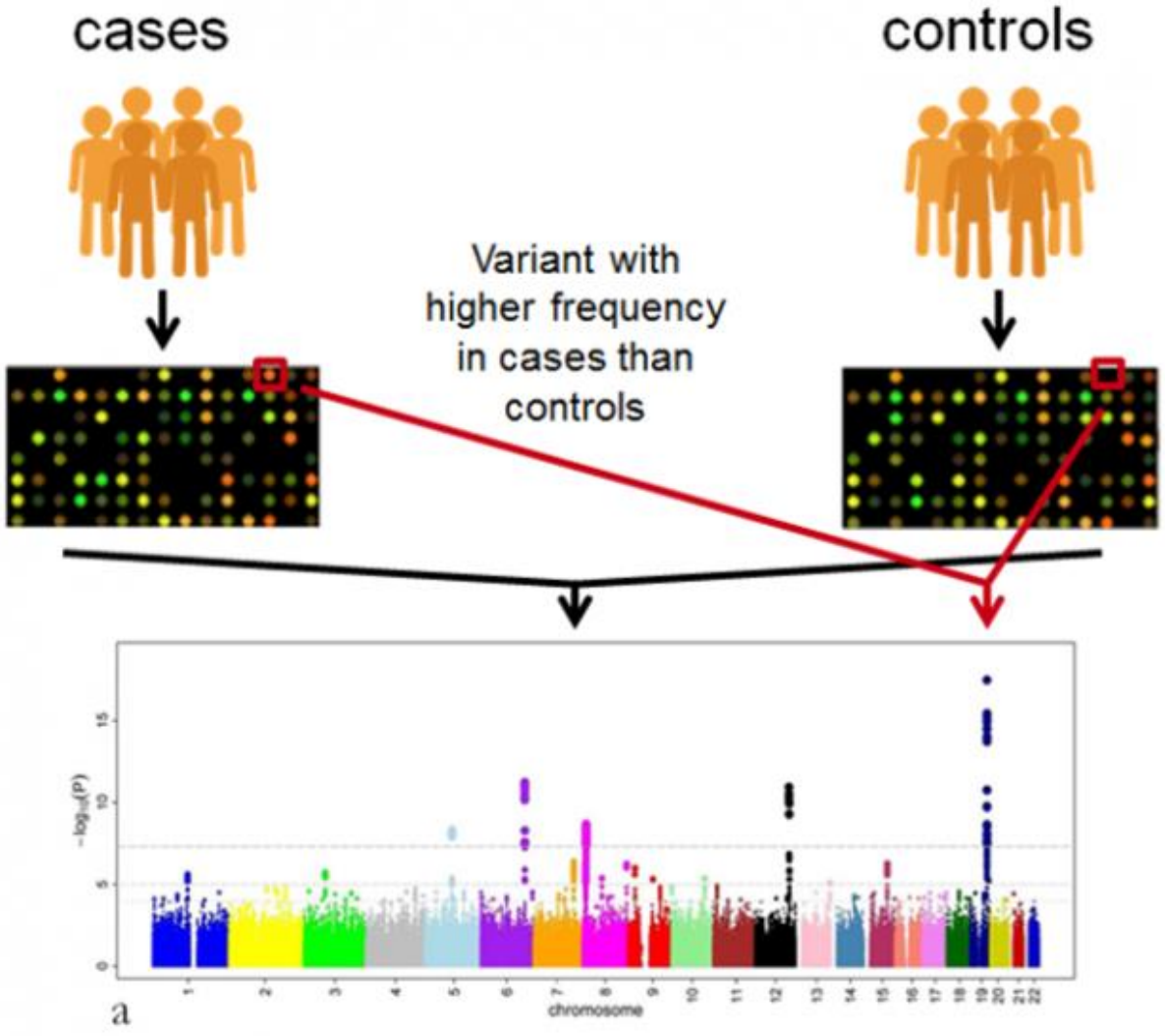


Genome-Wide Association Studies (GWAS)

Goal: Identify SNPs associated with traits or diseases

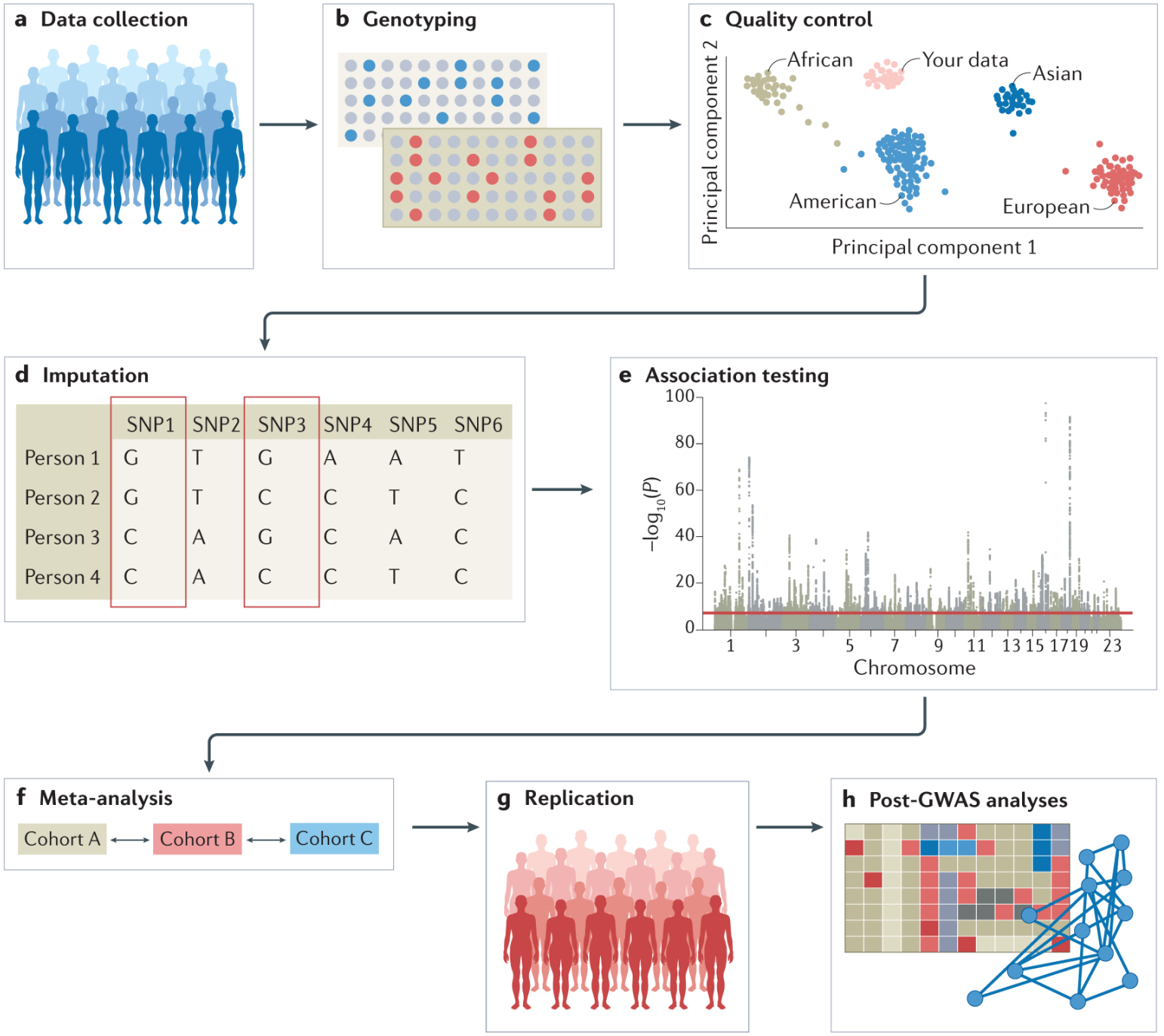
How: Compare SNP frequencies between case and control groups

Output: Manhattan plots highlighting significant associations



GWAS

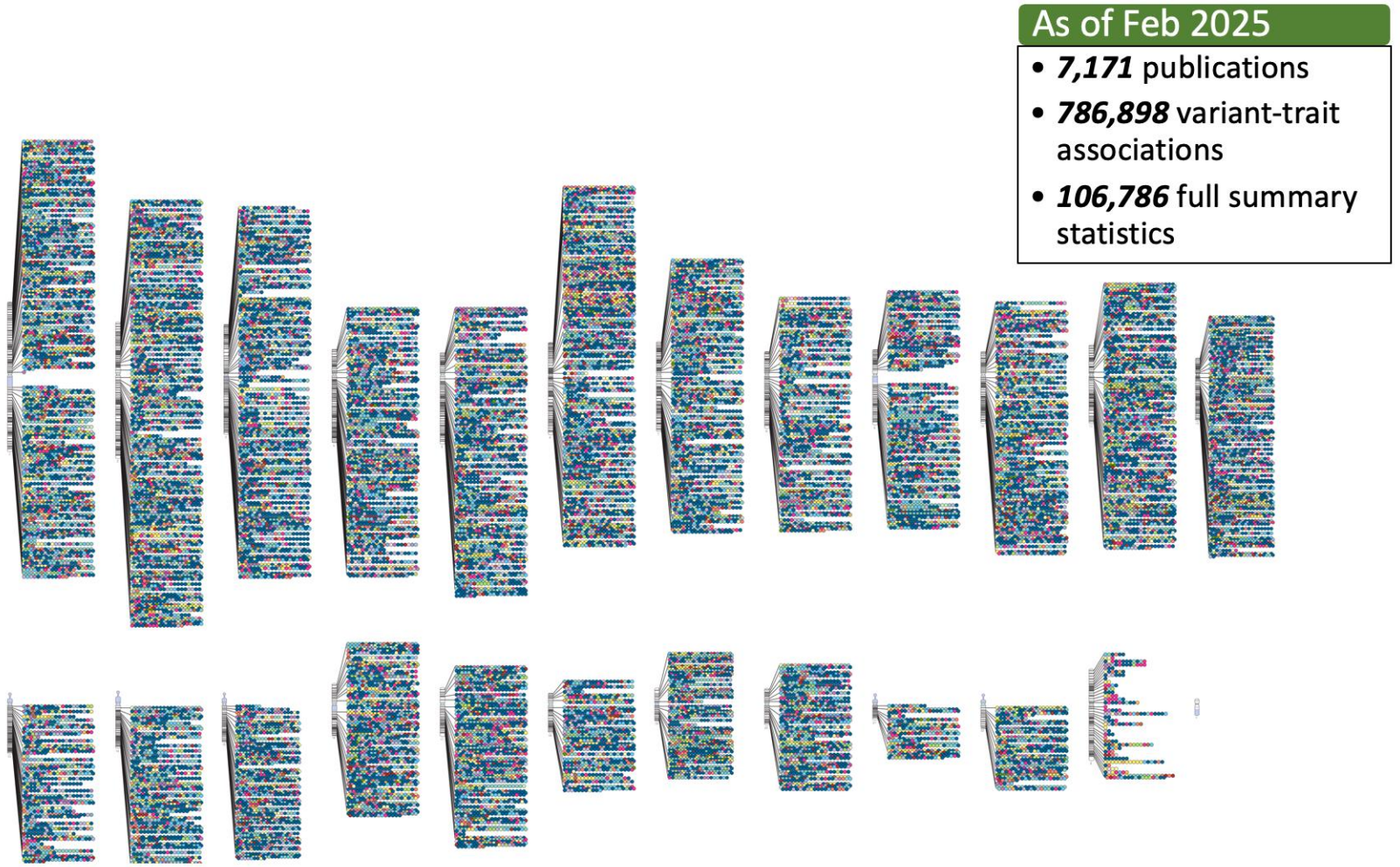
Standard GWAS workflow



GWAS catalog

Exploring SNP-trait associations

The NHGRI-EBI GWAS Catalog is a publicly available resource of Genome Wide Association Studies (GWAS) and their results.



References

https://www.illumina.com/techniques/popular-applications/genotyping/snp-snv-genotyping.html

https://online.stat.psu.edu/stat555/node/18/

https://www.ebi.ac.uk/training/online/courses/human-genetic-variation-introduction/

https://eu.idtdna.com/pages/education/decoded/article/genotyping-terms-to-know

https://www.ebi.ac.uk/training/online/courses/gwas-catalogue-exploring-snp-trait-associations/